心理学报 2025, Vol. 57, No. 2, 298−314 https://doi.org/10.3724/SP.J.1041.2025.0298
Acta Psychologica Sinica https://cstr.cn/32110.14.2025.0298

负面绩效反馈下员工绩效改进动机的人机比较*

王国轩[1] 龙立荣[1] 李绍龙[2] 孙 芳[3] 望家晴[1] 黄世英子[1]

([1]华中科技大学管理学院, 武汉 430074) ([2]武汉大学经济与管理学院, 武汉 430072)
([3]湖北经济学院工商管理学院, 武汉 430205)

摘 要 负面绩效反馈对员工学习和绩效提升具有重要意义, 然而其往往难以被员工所接受。随着人工智能技术(artificial intelligence, AI)逐渐应用于组织情境中, 探索 AI 提供负面绩效反馈对员工行为及态度的影响成为重要议题。采用 4 个递进式实验探索了 AI 与人类管理者提供负面绩效反馈对个体绩效改进动机的差异化影响及机制。实验 1~3 采取经典的虚假反馈的策略, 发现相较于人类管理者, AI 提供负面绩效反馈引发个体更高水平的绩效改进动机(实验 1)。并且, 在客观任务中, AI (较人类管理者)提供负面绩效反馈引发个体更高水平的绩效改进动机; 而在主观任务中, 结果则相反(实验 2)。此外, 个体对于负面绩效反馈的内部归因解释了上述关系发生的内在机制(实验 3)。实验 4 则采用相对真实的负面绩效反馈情境, 重复了先前 3 个实验的研究发现。该研究对于组织为何以及何时应用 AI 提供负面绩效反馈提供了一定的启示。

关键词 负面绩效反馈, 人工智能, 绩效改进动机, 内部归因, 任务类型

分类号 B849: C93

1 前言

负面绩效反馈是组织对未达到业绩期望的员工所给予的否定和批评(Cianci et al., 2010)。通常来说, 管理者提供负面绩效反馈的目的在于引导和激励员工的绩效表现(Lam et al., 2011; Podsakoff & Farh, 1989)。然而遗憾的是, 负面绩效反馈多引发员工焦虑、悲伤等负面情绪, 从而降低员工绩效(Audia & Locke, 2003; Kitz et al., 2023)。此外, 由于涉及人际沟通, 负面绩效反馈还会降低管理者与员工的关系质量(Ni & Zheng, 2024)。特别是在中国文化下, 考虑到人们的沟通方式较为含蓄, 管理者的负面绩效反馈使员工愧疚和尴尬, 继而损害工作积极性(耿紫珍 等, 2020)。盖勒普(Gallup)[1] 在 2019 年的调查也显示, 在对管理者负面绩效反馈产生消极情绪(失望、沮丧)后, 仅有 10.4%的员工会继续投入工作或改善绩效水平。综合看来, 传统由人类管理者提供负面绩效反馈的方式面临着较大的挑战(Kluger & Denisi, 1996; Xing et al., 2023)。

随着数智化技术的发展, 采用 AI 提供负面绩效反馈为组织带来新的机遇(Lee, 2018; Luo et al., 2021)。例如, 一款名为 Enaible 的算法通过远程监控员工的工作行为, 诊断员工表现不佳的原因, 并提供绩效改进建议。此外, Butterfly 等 AI 评估软件可以细致搜集员工的行为数据, 帮助员工及时改善表现(Tong et al., 2021)。那么, AI 相较人类管理者提供负面绩效反馈有哪些潜在优势? 研究发现 AI 有着强大的数据整合及分析能力, 且具备较少的主观意图(Garvey et al., 2023; Lee, 2018)。因此, 相较于人类管理者, AI 提供负面绩效反馈会更加“对事不对人” (Yalcin et al., 2022), 继而削弱传统人际互动中员工对负面绩效反馈的归因偏差(attribution bias) (例如, 将负面绩效反馈归因于管理者的偏见) (Xing et al., 2023), 使得员工更多关注自身的不足,

收稿日期: 2023-06-29

* 国家自然科学基金项目(72132001, 72272114, 72302082)资助。

通信作者: 李绍龙, E-mail: tli@whu.edu.cn

[1] “Why Employees Are Fed Up With Feedback” (gallup.com)

并增强绩效改进动机。

尽管已有文献初步表明, AI 和人类管理者在提供绩效反馈时可能呈现出不同的特征(Garvey et al., 2023; Yalcin et al., 2022), 员工在与 AI 或人类互动时也会产生差异化的反应(Tong et al., 2021), 但少有研究在人机提供负面绩效反馈情境下探讨员工的归因过程及后续反应, 因此本研究的目标包括: 首先, 基于人机比较的研究, 本研究拟探索人机负面绩效反馈(即由人类管理者或 AI 提供负面绩效反馈)的差异化影响效应。第二, 当前算法态度[2]的研究表明, 人类对 AI 的欣赏或厌恶取决于任务类型(Castelo et al., 2019)。比如在客观任务中, 相比于人类反馈, 个体更愿意接受准确性更强的 AI 反馈。因此研究拟探索任务类型(主观或客观任务)在人机负面绩效反馈中的边界效应。最后, 员工会对负面绩效反馈进行内部或外部归因, 从而决定是否改善绩效(Ilgen et al., 1979; Tolli & Schmidt, 2008)。基于此, 研究拟基于归因理论的视角, 探究人机提供负面绩效反馈产生差异化结果的机制。

1.1　人机负面绩效反馈对个体绩效改进动机的影响

传统人际场景中, 负面绩效反馈对员工消极的影响被广泛探讨, 例如, 减少学习动机(Xing et al., 2023)、自我效能感(Dimotakis et al., 2017), 降低员工目标设置以及绩效改进(Podsakoff & Farh, 1989), 或阻碍创造力(Kim & Kim, 2020)。上述负面影响大致通过三条路径解释: 人际破坏性、负面情绪和自我防御路径。首先, 员工可能将负面绩效反馈知觉为管理者的敌意, 致使学习和改善意愿下降(Cianci et al., 2010; Ni & Zheng, 2024)。第二, 在受挫、羞愧等负面情绪状态的影响下, 员工受到打击并较少关注绩效提升(Belschak & Den Hartog, 2009; Kim & Kim, 2020)。最后, 受自我防御的驱使, 负面绩效反馈减少员工内部归因, 导致绩效水平降低(Xing et al., 2023)。

人类与 AI 提供反馈的特征存在较大差异。由于依赖主观的经验与直觉, 人类管理者的反馈容易包含个人看法或偏见(蒋路远 等, 2022; Qin et al., 2023), 继而引发员工的消极情绪和防御反应(Ni & Zheng, 2024)。相反, AI 作为反馈提供者不容易出现认知疲劳和情绪失控, 并且由于具备强大的数据分析与预测能力, 使 AI 反馈更加客观与全面, 也较少被个体知觉为恶意或偏见(许丽颖 等, 2022; Qin et al., 2023)。值得说明的是, 人类与 AI 提供反馈的特征差异在负面情境中更为显著。比如, 面对 AI 提供的负面信息, AI 的客观性减少了个体对其蓄意意图的感知, 并增加信息接受度(Garvey et al., 2023)。相反, 囿于主观偏差, 人类的负面决策容易包含个人看法或主观判断, 从而降低接受度(Tong et al., 2021)。另外, 在负面事件中, 相对于人类, 个体更容易接受 AI 的决策。比如, 面对商品价格不公平时, 相较于人类, 消费者认为 AI 的决策基于大量客观数据生成, 进而产生更高水平的信任, 而人类销售员的决策可能存在主观局限性, 从而引发消费者较高的蓄意评估(宋晓兵, 何夏楠, 2020)。此外, 当个体受到监控时, 相较于人类监控, 算法监控被认为具有较低的主观判断与意志, 而更容易被接受(Raveendhran & Fast, 2021)。基于上述分析提出:

假设 1: 相较于人类管理者, AI 提供负面绩效反馈引发员工更高水平的绩效改进动机。

1.2　任务类型作为边界条件

在组织中, 与绩效关联的任务通常有主观与客观之分(Van Dijk & Kluger, 2011)。前者是基于个人观点或直觉的开放式或可解释任务(处理人际关系以及沟通等)。而后者是可量化的事实型任务(业绩分析、销量预测等) (Castelo et al., 2019)。任务类型是影响个体偏好人类或 AI 决策的关键因素。比如, 相较于人类, 用户认为算法基于客观销量数据提供的购买建议更为公平与中肯, 因此更愿意接受来自 AI 的建议(Helberger et al., 2020)。此外, 主观任务依赖人际互动能力, 需要通过直觉、经验和隐性知识等处理(Castelo et al., 2019), 而人类相较于 AI 在社会属性与主观属性方面更具优势。因此相比于 AI, 个体更信任人类管理者对主观任务提供的负面绩效反馈(Newman et al., 2020), 进而产生较高水平的绩效改进动机。相反, 客观任务具有可量化的特点(Castelo et al., 2019), 能够充分发挥 AI 负面绩效反馈在强大算力加持下的客观属性优势(Longoni et al., 2019; Tong et al., 2021), 使得员工更能接受反馈并增强绩效改进动机。据此提出:

假设 2: 人机负面绩效反馈与任务类型交互影响员工的绩效改进动机水平。具体而言, 在客观任

[2] 广义上来说, 算法可被分为脚本型(scripted-based)与机器学习型(machine-learning-based) (Lyytinen et al., 2021)。前者多属于数学概念, 后者则兼备人工智能的原理。由于算法是 AI 重要的驱动力, 因此本研究也关注机器学习类算法的文献。

务中, 相较于人类管理者, AI 提供负面绩效反馈会引发员工更高水平的绩效改进动机; 在主观任务中, 相较于 AI, 人类管理者提供负面绩效反馈会引发员工更高水平的绩效改进动机。

1.3 内部和外部归因的中介作用

归因理论(attribution theory; Heider, 1958)的因果控制点(locus of causality)视角将个体的归因风格区分为内部与外部归因。内部归因强调, 个体倾向寻找自身原因, 并相信当前的结果与个人因素(比如, 个人能力或性格特征等)相关; 相反, 外部归因则表示个体之所以出现某种行为或结果, 与所处的环境或运气等外部因素关联。

面对负面绩效反馈, 个体会识别反馈提供者的意图, 从而选择内部或外部归因(Audia & Locke, 2003)。具体来说, 当感知负面绩效反馈出于管理者的恶意时(例如, 打压, 伤害等), 个体倾向于外部归因。相反, 当负面绩效反馈传递出管理者帮助员工改善绩效的意图时, 个体则更多地内部归因(Ni & Zheng, 2024; Xing et al., 2023)。由于 AI 的决策依赖客观数据, 使 AI 具备更少的主观意图(Garvey et al., 2023)。因此, 相比于人类, AI 在负面情境中输出的决策具有更低的蓄意性与伤害性, 也更容易被接受。比如, 相比于人类歧视, 个体认为算法歧视具有更低的自由意志, 因此对其道德惩罚欲更少(许丽颖 等, 2022)。再如, 面对高于预期的价格, 个体认为 AI (较人类)的出价具备较低的主观意图并更愿意接受(Garvey et al., 2023)。总结来看, 相比于人类管理者, AI 基于客观数据与或已有事实的特性使负面绩效反馈更加客观, 且具有更少的主观意图(Tong et al., 2021), 从而提升员工的内部归因(Yalcin et al., 2022)。

此外, 研究发现, 经历挫折等负性事件会增强个体的成就动机(achievement motivation)。特别是对负面绩效反馈内部归因后, 员工会完善自身的行为表现, 以期达到更高的绩效表现来维护自尊水平(Weiner, 1985)。相反, 当员工将绩效反馈结果归因于自身无法控制的外部因素时, 可能会产生无力感并导致绩效改进动机下降(Harvey et al., 2014)。据此, 本研究认为, 当员工将负面反馈归因于自身因素(即内部归因)而非外部情境因素(即外部归因)时, 能够提高员工的绩效改进动机水平。因此提出:

假设 3: 内部与外部归因分别中介了人机负面绩效反馈对绩效改进动机水平的影响。具体而言, 相较于人类管理者, 员工对 AI 负面绩效反馈的内部归因水平更高、外部归因水平更低, 进而会产生更高水平的绩效改进动机。

最后, 由于客观任务可量化的特点, 相比于人类管理者, 客观任务下 AI 的数据整合与分析能力更强, 更容易获得个体的信任, 进而引发员工的内部归因(和减少外部归因), 并提高绩效改进动机水平。相反, 在主观任务中, 相比于 AI, 人类管理者拥有的人际沟通经验和互动能力能够更好地评估员工表现。因此, 在主观任务中, 相比于 AI, 人类管理者提供的负面绩效反馈会引发更高水平的内部归因(以及更低的外部归因), 从而增强个体的绩效改进动机(Castelo et al., 2019)。综上提出:

假设 4: 内部与外部归因分别中介了人机负面绩效反馈和任务类型对绩效改进动机的交互作用。具体而言, 在客观任务中, 相较于人类管理者, 员工对 AI 负面绩效反馈的内部归因水平更高、外部归因水平更低, 进而产生更高水平的绩效改进动机; 在主观任务中, 相较于 AI, 个体对人类管理者负面绩效反馈的内部归因水平更高、外部归因水平更低, 进而产生更高水平的绩效改进动机。

1.4 研究概览

本研究关注的主要问题是, AI 或人类管理者提供的负面绩效反馈是否会影响员工差异化的绩效改进动机, 并探讨任务类型是否在上述过程中发挥边界效应, 以及内部与外部归因的中介作用。本研究拟通过 4 个递进的实验对假设进行检验。具体来说, 为达到对负面绩效反馈内容更好的控制, 实验 1~3 采用虚假反馈的策略为个体提供负面绩效反馈(即被试收到内容完全相同的反馈)。其中, 实验 1 在 Credemo 平台上进行, 目的在于检验研究的假设 1, 即相较于人类管理者, 由 AI 提供的负面绩效反馈是否会导致员工产生更高水平的绩效改进动机。实验 2 在实验 1 检验主效应的基础上, 采用在问卷网定向招募不同行业与岗位员工的取样策略(即通过调查平台发布实验信息与要求, 招募愿意参与本研究的员工被试), 目的在于检验研究的假设 2, 并进一步探索任务类型的调节效应。此外, 实验 3 通过发送工作邮件这一更真实的绩效反馈形式为员工提供反馈信息, 并进一步检验内部与外部归因的中介效应(假设 3 与 4)。最后, 为进一步提高反馈的质量并增强其相对真实性, 实验 4 向个体提供相对而言更加真实的负面绩效反馈(基于个体任务表现的真实评估, 且更加具体和准确的反馈)。为提升研究结果的适用性, 本研究采用不同类型的 AI 代理

提供负面绩效反馈。具体来说, 实验 1 为嵌入型 AI (算法), 而实验 2~4 为机器人式 AI。

2　实验 1：AI 提供负面绩效反馈引发较高的绩效改进动机

实验 1 的目的是探究人机提供负面绩效反馈对员工绩效改进动机的差异化影响。

2.1　方法

2.1.1　被试

采用 G*Power 3.1 (Faul et al., 2007)计算本实验所需样本量。对于本实验适用的单因素方差分析, 取中等效应量 f = 0.25, 显著性水平 α = 0.05, 组别数为 2。事前检验显示, 为达到 80%的统计检验力至少需要 128 名被试。通过 Credamo 平台发布实验, 实时剔除 12 份注意力检查未通过、规律性作答或回答时间过短的样本, 并滚动采集, 最终得到了 128 份有效数据。其中, 女性 71 名(55.5%), 被试平均年龄为 33.95 岁(SD = 8.24)。参与实验的被试被随机分配到人类管理者或 AI 负面绩效反馈组, 其中人类管理者组 63 人, AI 组 65 人。所有被试均自愿参加实验并知情同意。通过注意力检查并完成实验的被试可获得一定报酬。

2.1.2　实验设计与程序

实验 1 采用单因素两水平的被试间设计：人机(人类管理者 vs. AI)负面绩效反馈。被试被随机分配到两个实验组别中的一组。首先, 考虑到自我效能感水平可能影响个体收到负面绩效反馈后的态度(Kluger & Denisi, 1996), 为控制这一无关变量, 要求被试先填写 10 题项的一般自我效能感问卷(Scholz et al., 2002), 采用 7 点量表计分, 从“1 = 非常不同意”到“7 = 非常同意” (实验 1 该测量的内部一致性系数为 0.91)。被试会阅读一段情境材料, 情境材料改编自 Tong 等(2021)的研究。描述了某公司呼叫中心电话销售员的日常工作情境, 并告知被试目前是该公司的一名电话销售, 公司为了评估呼叫中心电话销售员的工作表现, 设立了质量控制部门, 对销售员的服务电话进行录音分析, 并于每周固定时间通过公司内部绩效反馈系统提供绩效反馈。

以往研究表明, 个体可能会因为 AI 规则或原理的不透明性从而降低对 AI 的信任(Glikson & Woolley, 2020)。因此在 AI 组, 适当地为被试提供 AI 的属性信息, 被试会被告知(人机反馈的变化通过字体加粗显示): “**质量控制部门的人工智能系统小 ai (基于算法系统，由测评专家设计评价标准，人工智能学者和计算机专家开发的用于提供绩效反馈的程序)，通过对电话录音和销售量的数据分析为电话销售员提供专业的绩效反馈**”。在人类管理者组, 相应地告知被试: “**质量控制部门的销售经理小艾(经历过系统的绩效管理培训，具有专业知识和从业经验)，通过对电话录音和销售量的数据分析，为电话销售员提供专业的绩效反馈**” (图片材料见网络版附录)。随后, 为检验人机负面绩效反馈的启动, 要求被试填写操纵检验题目(“刚才是谁为您提供绩效反馈?”)。

接下来, 不同组别的被试会收到来自“销售经理小艾”或“人工智能系统小 ai”的负面绩效反馈:“你的工作表现低于部门平均水平, 你现在是部门绩效表现较低的员工之一, 希望你能持续改进” (Belschak & Den Hartog, 2009)。阅读完反馈信息后, 要求被试报告绩效改进动机。采用并改编来自 Wexley 等(1973)的两题项量表(实验 1~4 均采用该测量的题项): “收到绩效反馈后, 你多大程度上想在未来工作中达到更高的绩效目标?”和“收到绩效反馈后, 你多大程度上想在未来工作中提升绩效” (采用 7 点量表评分, 从“1 = 一点没有”到“7 = 非常”)。实验 1 该测量的内部一致性系数为 0.84。

考虑到 AI 组中被试可能对 AI 的熟悉度有所不同, 进而影响 AI 反馈后的绩效改进动机。为排除上述影响, 参考 Leo 和 Huh (2020)的研究, 要求被试报告对于 AI 的熟悉程度, 采用两题项测量(“请问在你的日常工作或生活中是否经常与人工智能打交道?”和“请问你对人工智能的工作原理和运行机制是否熟悉和了解?”; 从“1 = 一点不了解”到“7 = 非常了解”)。另外, 在材料阅读或问卷填写过程中会随机出现两道注意力检测题目(此题请选择“非常不同意”用于筛选未认真作答的被试)。最后, 被试报告了性别和年龄两项人口统计学信息。

2.2　结果

2.2.1　操纵检验

为检验人机提供负面绩效反馈的启动效果, 要求被试阅读实验材料后回忆负面绩效反馈的提供者“请回忆刚才是谁为你提供的绩效反馈”。经检查, 最终保留的 128 名被试全部回答正确。说明实验 1 对人机负面绩效反馈的操纵是成功的。

2.2.2　假设检验

独立样本 t 检验结果发现, 相比于人类管理者组(M = 4.94, SD = 1.38), AI 负面绩效反馈组(M = 5.49, SD = 1.18)中被试的绩效改进动机更强,

$t(126) = 2.38$, $p = 0.019$, Cohen's $d = 0.43$。为验证这一结果的稳健性, 首先将被试自我效能感作为控制变量, 进行单因素方差分析。结果表明, AI 组的绩效改进动机水平仍然高于人类管理者组, $F(1, 127) = 5.97$, $p = 0.016$, $\eta^2 = 0.046$。

其次, 为进一步排除被试性别(男性 = 1; 女性 = 2)和年龄可能对结果的影响, 分别对其进行相关分析和独立样本 t 检验。结果表明, 被试年龄与绩效改进动机并不相关($r = 0.07$, $p = 0.447$), 男性($M = 5.18$, $SD = 1.34$)和女性($M = 5.25$, $SD = 1.29$)的绩效改进动机也无显著差异, $t(126) = 0.34$, $p = 0.738$。最后, 为排除被试对 AI 的熟悉程度对结果可能的影响, 相关分析结果表明, AI 组被试的 AI 熟悉程度与绩效改进动机的相关不显著($r = 0.002$, $p = 0.990$)。

2.3 讨论

实验 1 在电话销售的绩效反馈场景中初步验证了研究假设 1, 即相对于人类管理者提供的负面绩效反馈, AI 提供的负面绩效反馈会促使个体产生更高水平的绩效改进动机。为验证实验 1 结果的稳健性, 并进一步检验假设 2, 实验 2 定向招募企业的在职员工, 并改变绩效反馈的场景(新员工入职培训场景), 拟验证任务类型与人机负面绩效反馈影响员工绩效改进动机的交互作用。此外, 在操纵人机反馈时, 实验 1 以嵌入式的算法作为 AI 组的实验材料。考虑到机器人式 AI (robotic AI)未来可能与员工共事并进入组织(Yam et al., 2023), 以及其通常在人机交互过程中引发更高水平的信任与体验感(Glikson & Woolley, 2020), 实验 2~4 采用机器人式 AI 作为 AI 组的实验材料, 并为个体提供负面绩效反馈。

3 实验 2: 任务类型作为边界条件

实验 2 定向招募企业的在职人员, 拟进一步探讨人机负面绩效反馈和任务类型是否交互影响员工绩效改进动机。此外, 为进一步提升实验材料的严谨性, 实验 2 对实验材料进行了预测试。

3.1 方法

3.1.1 实验材料预测试

在正式进行实验 2 之前, 为了检验本研究自编的任务类型(主观 vs. 客观)与人机(人类管理者 vs. AI)反馈刺激图片的可靠性。在 Credamo 平台招募 60 名被试(男性 29 名, 女性 31 名, 被试平均年龄为 29.07 岁($SD = 7.97$), 对实验材料展开预测试。被试被随机分配到主观或客观任务情境中, 完成三道任务题目(每题作答不少于 100 字)。其中, 主观任务中的三道题目分别包含冲突化解, 突发事件应对以及人际沟通等组织中常面临的问题; 而客观任务的三道题目分别包含人员排序, 方案计算以及销量预测三方面内容(详见网络版附录)。此外, 被试需要对主观任务和客观任务的任务客观性(您在多大程度上认为上述任务是客观任务; 1 = 一点也不, 5 = 很大程度上)、任务难度进行评估(请您评估上述任务的难易程度; 1 = 一点不难, 5 = 非常困难)。独立样本 t 检验结果显示, 客观任务($M = 4.10$, $SD = 0.80$)的任务客观性显著高于主观任务($M = 1.53$, $SD = 0.82$), $t(58) = 12.25$, $p < 0.001$, Cohen's $d = 0.80$; 此外, t 检验结果表明, 两类任务的难度不存在显著差异, $t(58) = 0.26$, $p = 0.25$。同时, 为检验两种任务类型是否贴近真实的工作场景, 采用 5 题项量表(Fields et al., 2023), 并要求被试进行表面效度(face validity)评分, 代表题项如:"您在多大程度上认为, 本次测试的实际内容是与日常工作明显相关的", 从"1 = 完全没有"到"5 = 完全是"。结果显示, 两种任务的表面效度均较高, 主观任务的表面效度均值为 4.33, 而客观任务为 3.92。因此可知, 实验任务的设置较为合理且比较符合现实的工作场景。

此外, 参考 Garvey 等(2023)的实验材料, 检验 AI 和人类管理者形象的图片在面孔吸引力和诡异性方面的差异(图片材料见网络版附录)。独立样本 t 检验结果显示, 两张图片在面孔吸引力上不存在显著差异, $t(59) = 0.24$, $p = 0.59$。人工智能图片仅存在轻微的诡异度($M = 1.47$, $SD = 0.57$)。因此可知, 实验 2 对刺激图片的选择较为合理。

3.1.2 被试

采用 G*Power 3.1 软件(Faul et al., 2007)计算实验所需的样本量。对于本实验适用的双因素方差分析, 取中等效应量 $f = 0.25$, 显著性水平 $\alpha = 0.05$, 组数为 4。事前检验表明, 要达到 85%的统计检验力至少需要 146 名被试。实验 2 委托问卷网发布实验信息, 定向招募在职员工被试参与本实验, 并在平台导入实验相关材料, 开展线上行为实验。考虑到可能会出现未完成或回答无效的数据, 实验二招募了 168 名在职企业人员。剔除 8 份没有通过注意力检查、存在规律性回答或作答时间异常的数据后, 最终得到了 160 份有效数据。其中, 女性 61 名(38.1%), 被试平均年龄为 33.29 岁($SD = 4.98$)。被试主要来自制造、软件、金融、教育及快消品行业,

主要从事管理(55 人, 占 34.4%)、生产运营(37 人, 占 23.1%)、技术研发(10 人, 占 6.3%)、市场营销(25 人, 占 15.6%)、产品设计(33 人, 占 20.6%)等岗位的工作。所有被试均自愿参加实验并知情同意。通过注意力检查并完成实验的被试可获得相应报酬。

3.1.3　实验设计与程序

实验 2 采用双因素被试间设计: 2 (人机负面绩效反馈：人类管理者 vs. AI) × 2 (任务类型：主观 vs.客观)。参与实验的被试被随机分配到 4 个实验组别中的其中一组。

正式实验前, 同实验 1, 要求被试填写一般自我效能感问卷(Scholz et al., 2002), 并作为研究的控制变量(实验 2 该测量的内部一致性系数为 0.89)。被试随后会阅读一段情境材料, 告知被试是某日用品股份有限公司的一名新入职员工。在一次部门月度工作总结结束后, 部门为包括被试在内的 5 名新员工安排了职业能力的测试, 以更好地制定个性化的培训计划, 并为后续的岗位安排提供基础。接下来, 告知被试以下是测试的其中一道代表性例题。在主观任务组, 被试需要完成一道职场中化解人际冲突的题目; 在客观任务组, 被试需要完成一道预测未来产品销量的题目(见网络版附录), 为确保被试认真作答并融入情景, 要求被试每题作答不少于 100 字。

完成测试题目后, 告知被试 2 分钟后会收到对于测试的反馈。在人类管理者组, 告知被试(加粗字展示了操纵人机反馈的差异): “公司邀请了人力资源部门测评专员王亮对你的表现进行评估反馈。**测评专员王亮(经过系统的职业能力测评培训, 具有专业知识且经验丰富的测评专家)会阅读你的回答, 评估你的作答质量, 并进行统计排名, 对你本次测试完成情况进行评估反馈**”, 并展示王亮形象的图片(见网络版附录)。在 AI 组, 被试相应被告知: “公司通过人力资源测评中心开发设计的人工智能评估助手小 ai, 对你的表现进行评估反馈。**人工智能评估助手小 ai, 会基于算法系统(该算法系统是基于测评专家设计的评价标准, 由人工智能学者和计算机专家开发的程序)对你的答案自动进行识别, 评估你的作答质量, 并进行统计排名, 对你本次测试完成情况进行评估反馈**” (相应展示小 ai, 见网络版附录)。随后, 要求被试填写操纵检验题目(例如“请问为您提供测试反馈的是”, 1 = 测评专员王亮, 7 = AI 评估助手小 ai 以及“请对您刚才完成的测试题目的客观性进行评分”, 1 = 非常主观, 7 = 非常客观)。

约 2 分钟后, 被试会收到来自“测评专员王亮或人工智能测评助手小 ai”的负面绩效反馈: “在本次测试中, 你的表现低于 80%的同事, 位于后 20%, 表现有待提升” (Kim & Kim, 2020)。最后, 要求被试填写绩效改进动机(实验 2 该测量的内部一致性系数为 0.84), 并报告性别、年龄、行业与岗位这 4 项人口统计学信息。AI 组被试需要填写对 AI 的熟悉程度。问卷填答过程中会随机出现两道注意力检测题目(此题请选择“非常不同意”)用于筛选未认真作答的被试。

3.2　结果

3.2.1　操纵检验

首先, 为检验人机提供负面绩效反馈的启动效果。要求被试阅读实验材料后回忆负面绩效反馈的提供者“请回忆刚才是谁为您提供的绩效反馈” (1 = 测评专员王亮, 7 = AI 评估助手小 ai)。结果显示, AI 负面绩效反馈组(M = 6.24, SD = 1.05)的评分显著高于人类管理者组(M = 1.95, SD = 1.11), t (158) = 25.11, p < 0.001, Cohen’s d = 0.86。说明实验 2 对人机负面绩效反馈的操纵成功。

其次, 为检验任务类型操纵有效性, 要求被试回答, “你认为刚才所做测试例题的客观性” (1 = 非常主观, 7 = 非常客观)。结果表明, 客观任务组(M = 5.96, SD = 0.79)的任务客观性评分显著高于主观任务组(M = 2.33, SD = 0.87), t(158) = 27.77, p < 0.001, Cohen’s d = 0.83。说明任务类型的操纵成功。

3.2.2　假设检验

其次, 独立样本 t 检验结果发现, 相较于人类管理者组(M = 5.36, SD = 1.13), AI 提供负面绩效反馈组(M = 5.67, SD = 0.79)被试的绩效改进动机更强, t (158) = 2.00, p = 0.048, Cohen’s d = 0.46。将被试自我效能感作为控制变量, 进行单因素方差分析。结果表明, AI 组的绩效改进动机水平仍然高于人类管理者组, F(1, 157) = 4.64, p = 0.033, η^2 = 0.029。为排除被试对 AI 的熟悉程度对结果可能的影响, 相关分析结果表明, AI 组被试的 AI 熟悉程度与绩效改进动机的相关不显著(r = 0.10, p = 0.40)。综上, 研究假设 1 再次得到验证。

最后, 检验任务类型是否能够作为边界条件。双因素方差分析结果表明, 人机负面绩效反馈和任务类型对个体的绩效改进动机有显著的交互影响, F(1, 156) = 39.65, p < 0.001, η^2 = 0.203。简单效应分析发现(如图 1), 在客观任务组, AI 负面绩效反馈

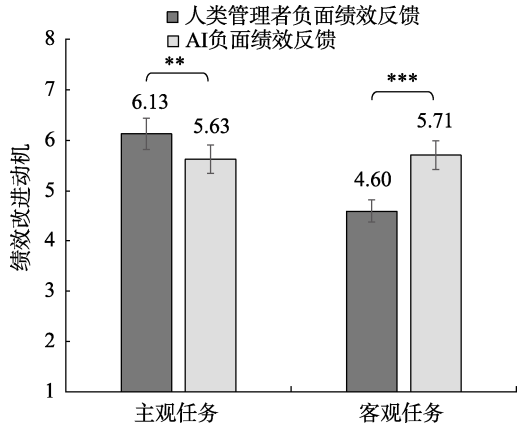

图 1 人机负面绩效反馈与任务类型对绩效改进动机的交互作用图

注：$^{**}p < 0.01$, $^{***}p < 0.001$.

组的绩效改进动机水平(M = 5.71, SD = 0.66)显著高于人类管理者组(M = 4.60, SD = 1.01), F (1, 156) = 37.75, $p < 0.001$, $\eta^2 = 0.195$。在主观任务组, 人类管理者组的负面反馈绩效改进动机水平(M = 6.13, SD = 0.59)显著高于 AI 组(M = 5.63, SD = 0.90), F(1, 156) = 7.63, $p = 0.006$, $\eta^2 = 0.047$。为检验上述结果的稳健性, 将被试自我效能感作为协变量进行控制, 结果发现, 人机负面绩效反馈与任务类型的交互作用依然显著, F(1, 155) = 40.58, $p < 0.001$, $\eta^2 = 0.207$。假设 2 得到验证。

3.3 讨论

实验 2 基于新入职员工绩效反馈的场景, 重复实验 1 研究发现的基础上, 还进一步发现了人机负面绩效反馈与任务类型对个体绩效改进动机的交互作用, 从而验证了研究假设 2。实验 2 较实验 1 主要有两点改善。首先, 参考以往绩效反馈的研究(Cianci et al., 2010), 在被试执行实验任务后间隔一定时间后提供绩效反馈, 以增强绩效反馈的真实性。其次, 实验 1 采用的是评价型反馈(例如, “你现在是部门绩效表现较低的员工之一”), 而实验 2 采用了客观型反馈(例如, “在本次测试中, 你的表现低于 80%的同事, 位于后 20%”), 这使负面绩效反馈的内容更加客观, 也有利于降低被试的负面情绪反应(Kim & Kim, 2020)。

但是, 实验 2 仍存在一些不足。比如, 虽然实验 2 通过任务执行−延迟反馈的设计在一定程度上提升了反馈的真实性, 但任务与绩效反馈的间隔时间以及绩效反馈的呈现方式仍然有待改进。鉴于此, 实验 3 在被试完成任务间隔 20 分钟后, 通过邮件这种更真实的形式发送绩效反馈, 并进一步检验内部与外部归因在人机负面绩效反馈影响员工绩效改进动机发挥的中介作用。

4 实验 3：内部与外部归因的中介作用

实验 3 目的是采用邮件绩效反馈这种更真实的形式, 进一步检验实验 1 和实验 2 实验结果的稳健性, 以及探讨内部和外部归因发挥的中介作用。

4.1 方法

4.1.1 被试

采用 G*Power 3.1 软件(Faul et al., 2007)计算本实验所需样本量。对于本实验适用的双因素方差分析, 取中等效应量 f = 0.25, 显著性水平 α = 0.05, 组别数为 4。事前分析显示, 要达到 85%的统计检验力至少需要 146 名被试。类似于实验 2, 实验 3 委托问卷网发布实验信息, 定向招募在职员工被试参与本实验。考虑到可能会出现少量未完成或填答无效的数据, 实验 3 共招募了 160 名在职员工参加实验。所有被试均自愿参加实验并知情同意, 通过注意力检测并完成实验任务的被试可获得相应报酬。

剔除没有通过注意力测试, 未完成作答或作答无效的被试 10 名, 实验 3 最终有效样本为 150。其中, 女性 86 名(57.3%), 被试平均年龄为 29.70 岁(SD = 4.97), 平均工龄为 6.17 年(SD = 4.23)。参与实验的被试来自互联网、建筑业、制造业、信息通信业、商品销售业、教育、医疗、金融和服务这 9 个行业。从工作岗位上来看, 参与实验的员工包含管理类 52 人, 占比 34.87%; 运营类 29 人, 占比 19.40%; 技术类 19 人, 占比 12.50%; 营销类 26 人, 占比 17.43%; 创意设计类 24, 占比 15.79% 。

4.1.2 实验设计与程序

实验 3 采用双因素组间设计: 2 (人机负面绩效反馈：人类管理者 vs. AI) × 2 (任务类型：主观 vs. 客观)。被试被随机分配到 4 个实验组别中的一组。

告知被试即将参与一场职业能力竞赛。事先收集参与者的工作邮箱以便后期发送对应的绩效反馈。竞赛分为两个阶段(竞赛与绩效反馈)。在竞赛阶段, 被试需要根据要求完成竞赛中的题目。首先, 要求被试填写自我效能感问卷以作为研究的控制变量(实验 3 该测量的内部一致性系数为 0.91)。接下来, 采用实验 2 的材料预测试题目, 被试依据组别完成三道主观或客观型竞赛题目(具体题目见网络版附录), 为确保被试认真作答并融入情景, 要求被试每题作答不少于 100 字。完成竞赛任务后,

被试被告知某高校的职业测评中心将负责对参与者的竞赛表现进行评估和反馈。在人类管理者组, 被试被告知(人机反馈的操纵差异通过加粗的字体呈现): “为了评估您在本次职业能力竞赛中的表现, 我们邀请了某高校测评中心的负责人王亮, 对您的表现进行评估反馈。**测评中心负责人王亮(经过系统的职业能力测评培训, 具有专业知识且经验丰富的测评专家), 会阅读您的回答, 评估您的作答质量, 并进行统计排名, 对您本次职业能力竞赛结果进行评估反馈**”。并配有王亮形象的图片(见网络版附录)。在 AI 组, 被试会被告知:“为了评估您在本次职业能力竞赛中的表现, 我们将使用某大学测评中心最新引进的人工智能测评助手小 ai, 对您的表现进行评估反馈。**人工智能测评助手小 ai, 基于算法系统(该算法系统是基于测评专家设计的评价标准, 由人工智能学者和计算机专家开发的程序)对您的答案自动进行识别和分析, 评估您的作答质量, 并进行统计排名, 对您本次职业能力竞赛结果进行评估反馈**” (相应展示小 ai, 见网络版附录)。期间, 要求被试回答操纵检验题目(“请问为您提供竞赛反馈的是”, 1 = 测评中心负责人王亮, 7 = 测评中心 AI 助手小 ai; 以及“请对您刚才完成的竞赛题目的客观性进行评分”, 1 = 非常主观, 7 = 非常客观)。随后, 被试被告知“由于需要等待和评估其他参与者的表现并进行最终排名, 绩效反馈需要约 20 分钟, 最终的竞赛结果和测评问卷链接会通过您的电子邮箱发送给您”。为减弱被试在等待反馈过程中可能出现的注意力缺失问题对实验结果的干扰, 要求所有被试观看一段时长约 20 分钟的某高校测评中心的介绍视频。

在绩效反馈阶段, 实验人员通过提前更换好域名的电子邮箱(比如, 某高校测评中心负责人王亮或 AI 测评助手小 ai), 向先前收集好的被试工作邮箱发送职业竞赛的反馈结果。为加强对人机负面绩效反馈的操纵, 人类管理者组的被试会收到: “您好! 在本次职业能力竞赛中, 您的表现低于 82%的参与者, 位于后 18%” (Kim & Kim, 2020)。而 AI 组除了为被试提供与人类管理者组相同内容的负面绩效反馈外, 还会在邮件的文末备注: “此邮件由人工智能助手自动发送, 请勿回复”。随后, 提示被试填答附于邮件中的第二阶段问卷, 包括回忆并简述绩效反馈内容(确保被试基于反馈信息填答后续题目); 对于 AI 的熟悉程度; 绩效改进动机(实验 3 该测量的内部一致性系数为 0.84); 内部与外部归因的测量选取 Russell (1982)开发的 6 题项量表来测量。3 题项用于测量内部归因, 代表题项如: “您在多大程度上认为, 反馈者提供的测评反馈结果是基于你个人的努力而产生”, 该测量内部一致性系数为 0.78; 外部归因代表题项如“您在多大程度上认为, 反馈者提供的测评反馈结果是基于环境因素而产生(比如题目很难)”, 该测量内部一致性系数为 0.74。采用 7 点评分法, 从“1 为一点也没有”, 到“7 为很大程度上”。

接下来, 已有研究发现, 个体对人类或 AI 绩效反馈的准确性(Tong et al., 2021)和公平感知(Newman et al., 2020)可能存在差异。比如, 由于 AI 本质上是一套数据驱动的程序模型, 其客观和无偏性更高, AI (较人类管理者)提供负面绩效反馈可能引发个体较高的准确性或公平感知(蒋路远 等, 2022), 从而差异化影响绩效改进动机。为排除上述两条替代解释机制, 要求被试填写反馈准确性(Brett & Atwater, 2001)和公平感量表(Chory & Westerman, 2009)。反馈准确性的测量包含 2 个题项(“您在多大程度上认为您收到的反馈是对您表现的准确评估”; “您在多大程度上相信您收到的反馈是正确的”。采用 7 点量表评分, 从“1 = 一点没有”到“7 = 非常”, 该测量的内部一致性系数为 0.89)。公平感知包含 6 个题项(代表题项如“我认为反馈者给我提供的反馈是: 1 = 不公正的; 7 = 公正的”, “我认为反馈者给我提供的反馈是:1 = 有偏见的; 7 = 中立的”, 该测量的内部一致性系数为 0.97)。最后, 被试报告性别、年龄、工龄、行业以及岗位这 5 项人口学变量。问卷填答过程中会随机出现两道注意力检测题目(此题请选择“非常不同意”)用于筛选未认真作答的被试。

4.2 结果

4.2.1 操纵检验

首先, 为检验人机提供负面绩效反馈的启动效果。要求被试阅读实验材料后回忆负面绩效反馈的提供者“请回忆刚才是谁为您提供的绩效反馈” (1 = 测评中心负责人王亮, 7 = 测评中心 AI 助手小 ai)。结果显示, AI 负面绩效反馈组(M = 5.26, SD = 1.47)的评分显著高于人类管理者组(M = 2.49, SD = 0.86), $t(148) = 13.10$, $p < 0.001$, Cohen's d = 0.71。说明实验 3 对人机负面绩效反馈的操纵成功。

为检验任务类型的启动效果, 要求被试评价实验任务的客观性。结果表明, 客观任务组(M = 5.54, SD = 0.98)的任务客观性评分显著高于主观任务组(M = 3.41, SD = 1.95), $t(148) = 8.23$, $p < 0.001$,

Cohen's d = 0.63。表明任务类型的操纵成功。

4.2.2 假设检验

为检验研究假设 1。独立样本 t 检验结果表明, 相较于人类管理者(M = 4.66, SD = 1.10), AI 的负面绩效反馈(M = 5.06, SD = 1.21)能够引发更高水平的绩效改进动机, t(148) = 2.10, p = 0.037, Cohen's d = 0.44。将自我效能感作为协变量控制, 依然发现相较于人类管理者, AI 提供负面绩效反馈会导致更高水平的绩效改进动机, F(1, 147) = 4.05, p = 0.046, η^2 = 0.027。假设 1 得到验证。其次, 为排除 AI 熟悉程度对结果的影响, 相关分析发现, AI 组被试对 AI 的熟悉度与其绩效改进动机并不相关(r = 0.06, p = 0.61)。为检验研究假设 2, 双因素方差分析结果发现(如图 2), 人机负面绩效反馈和任务类型对个体的绩效改进动机具有显著的交互作用, F(1, 146) = 20.00, p < 0.001, η^2 = 0.120。且简单效应分析发现, 在客观任务组, AI 负面反馈组的绩效改进动机水平(M = 5.60, SD = 0.88)显著高于人类管理者组(M = 4.35, SD = 0.92), F(1, 146) = 24.47, p < 0.001, η^2 = 0.144。在主观任务组, 人类管理者负面反馈组的绩效改进动机水平(M = 4.92, SD = 1.17)(边缘)显著高于 AI 组(M = 4.57, SD = 1.26), F(1, 146) = 3.00, p = 0.085 < 0.10, η^2 = 0.02。为检验上述结果的稳健性, 将被试自我效能感作为协变量进行控制, 结果发现, 人机负面绩效反馈与任务类型的交互作用依然显著, F(1, 145) = 22.79, p < 0.001, η^2 = 0.136。因此, 研究的假设 2 再次得到了验证。

为检验内部与外部归因的中介效应, 选择 PROCESS 插件的模型 4, Bootsrap 为 2000。结果表明, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应显著, 且中介效应的指标值为 0.17, 95%的 CI 为[0.015, 0.350], 区间不包含 0。此外, 外部归因在人机负面绩效反馈到绩效改进动机的中介效应指标值为 0.08, 95%的 CI 为[−0.04, 0.21], 区间包含 0, 说明外部归因在人机负面绩效反馈到绩效改进动机的间接效应不显著。研究假设 3 得到了部分验证。

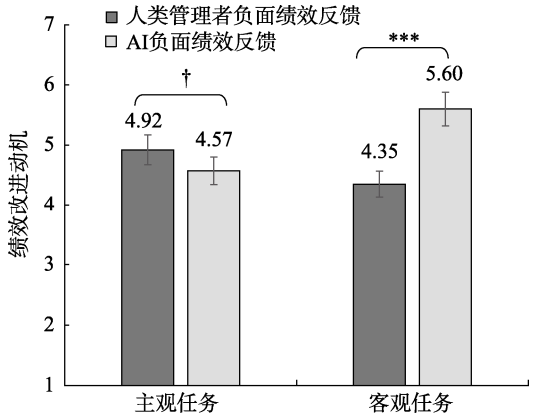

图 2 人机负面绩效反馈与任务类型对绩效改进动机的交互作用图

注: $^{\dagger}p$ < 0.10, $^{***}p$ < 0.001.

此外, 本研究还检验了内部与外部归因在人机负面绩效反馈与任务类型对绩效改进动机交互效应中的间接效应。选择 PROCESS 插件的模型 8, 设定样本量为 2000。结果表明, 在主观任务中, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应显著, 95%的 CI 为[−0.40, −0.03]; 在客观任务中, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应同样显著, 95%的 CI 为[0.20, 0.84]。且在两种不同类型的任务下, 有调节的间接效应的差值显著, 指标值为 0.67, 95%的 CI 为[0.277, 1.178]。说明有调节的中介效应显著; 而外部归因有调节的间接效应指标值为 0.12, 95%的 CI 为[−0.065, 0.375]。区间包含 0, 因此研究假设 4 也得到了部分验证。

为检验反馈准确性和公平感是否能作为替代性解释的机制。首先, 独立样本 t 检验发现, 人类管理者(M = 3.78, SD = 1.63)或 AI 提供负面绩效反馈(M = 3.80, SD = 1.60)在准确性水平上并无差异, t(148) = 0.07, p = 0.94。但是, AI 提供负面绩效反馈(M = 5.31, SD = 1.35)比人类管理者(M = 4.84, SD = 1.50)更公平, t(148) = 2.00, p = 0.047, Cohen's d = 0.32。此外, 人机负面绩效反馈与任务类型对反馈准确性(和公平感)的交互作用均不显著, F(1, 146) = 0.12, p = 0.73; F(1, 146) = 0.58, p = 0.45。最后, 反馈准确性(95%的 CI 为[−0.082, 0.065])和公平感(95%的 CI 为[−0.06, 0.24])在人机负面绩效反馈与绩效改进动机之间的间接效应也均不显著。综上, 实验 3 排除了反馈准确性与公平感这两条替代的解释机制。

4.3 讨论

实验 3 以更加贴近组织真实反馈的方式, 检验了实验 1 和实验 2 结果的稳健性, 并进一步发现了内部归因发挥的中介作用。这为不同任务情境下, AI 和人类管理者提供负面绩效反馈对员工绩效改进动机的差异化影响提供了一个良好的解释机制。不仅如此, 实验 3 还排除了反馈准确性与公平感的替代解释机制。

实验 3 未能发现外部归因的中介作用, 可能的

原因在于, 个体出于自尊维护与自我防御的目的, 无论人类管理者抑或 AI 提供负面绩效反馈, 可能都倾向于表现一定水平的外部归因(Hareli & Hess, 2008)。比如实验 3 中, 人类管理者组(M = 3.60, SD = 1.27)与 AI 负面绩效反馈组(M = 3.92, SD = 1.33)的被试均有一定程度的外部归因, 但差异并不显著, $t(148) = 1.52$, $p = 0.13$。与实验 3 结果一致的是, Yalcin 等(2022)也发现在不利性决策情境中(例如遭到公司拒绝), 消费者对来自人类或 AI 客服反馈的外部归因差异不显著。

此外, 实验 1~3 采用了绩效反馈研究中常用的虚假反馈范式(Cianci et al., 2010; Kim & Kim, 2020), 其优点在于控制被试间反馈内容的一致性。但由于没有评估被试的真实任务表现, 导致个体可能对负面绩效反馈的准确性、真实性感知较低。为解决上述问题并进一步提升绩效反馈的质量, 实验 4 拟在相对真实的反馈场景下为个体提供更为具体和个性化的绩效反馈, 从而再次验证本研究的整体模型。

5　实验 4：真实反馈下对假设模型的再验证

实验 4 拟基于个体真实的任务表现为其提供具体的、个性化的绩效反馈。具体来说, 除提供结果性的反馈外(展示排名, 成绩等结果), 实验 4 加入了过程性的反馈, 比如解释任务目的, 提供针对任务表现的鼓励, 或问题诊断等具体信息(Goodman & Wood, 2004), 从而加强负面绩效反馈的真实性和质量。

5.1　方法

5.1.1　被试

采用 G*Power 3.1 软件(Faul et al., 2007)计算本实验所需样本量。对于本实验适用的双因素方差分析, 取中等效应量 f = 0.25, 显著性水平 α = 0.05, 组别数为 4。事前分析显示, 要达到 85%的统计检验力至少需要 146 名被试。类似于实验 2 和 3, 实验 4 委托问卷网发布实验信息, 定向招募在职员工被试。考虑到可能会出现少量未完成或填答无效的数据, 实验 4 共招募了 166 名在职员工参加实验。所有被试均自愿参加实验并知情同意。通过注意力检测并完成实验任务的被试可获得相应报酬。

剔除没有通过注意力测试, 未完成作答或作答无效的被试 6 名。实验 4 最终有效样本为 160。其中, 女性 65 名(40.60%), 被试平均年龄为 29.54 岁(SD = 6.07)。91.3%的被试有过本科及以上的教育经历。参与实验的被试主要来自制造业、软件业、商务服务业、金融业、科学研究与教育业等 5 个行业。从工作岗位来看, 参与实验的员工包含技术研发类 50 人, 占比 31.30%; 管理类 34 人, 占比 21.30%; 生产与运营类 31 人, 占比 19.40%; 市场营销类 31 人, 占比 19.40%; 产品设计类 14 人, 占比 8.80% 。

5.1.2　实验设计与程序

实验 4 采用双因素组间设计: 2 (人机负面绩效反馈：人类管理者 vs. AI) × 2 (任务类型：主观 vs. 客观)。被试被随机分配到 4 个实验组别中的一组。

在正式实验开始前进行两项实验的准备工作。首先, 对 5 名主试人员进行培训, 使他们清晰并熟练地掌握各题目的作答要点或评价标准。其次, 事先编辑好负面绩效反馈的模版, 在正式作答时依据被试的表现进行个性化的更改, 从而控制反馈的内容并减少提供反馈所需的时间。在正式实验阶段, 第一, 告知被试是某涂料公司的一名主管, 即将参与一场针对企业中层管理人员的管理能力测评, 为后续的培训与学习提供参考依据。随后, 将测评分为两个阶段(任务以及绩效反馈)。在测评任务阶段, 被试需要完成以下实验步骤：第一, 要求被试填写自我效能感问卷以作为研究的控制变量(实验 4 该测量的内部一致性系数为 0.96)。第二, 与实验 2~3 类似, 不同任务组别的被试需要相应完成 2 道主观或客观任务(见网络版附录), 要求每题的作答不得少于 100 字。第三, 完成测评任务后, 被试被告知公司将会对其测评表现进行专业评估和反馈。在人类管理者组(人机反馈操纵的差异通过加粗字体展示): “为了评估您在本次管理能力中的表现, 我们邀请公司管理能力测评专员王亮, 对您的表现进行评估反馈。**测评专员王亮(经过系统的职业能力测评培训，具有专业知识且经验丰富的测评专家), 会阅读您的回答，评估您的作答质量，并进行统计排名，对您本次管理能力测评的结果进行评估反馈**”, 并搭配一张王亮形象的图片(见网络版附录)。在 AI 组, 则对应描述为: “为了评估您在本次职业能力竞赛中的表现, 我们将使用公司人事部最新引进的人工智能(AI)测评助手小 ai, 对您的表现进行评估反馈。**人工智能测评助手小 ai，基于算法系统(该算法系统是基于测评专家设计的评价标准，由人工智能学者和计算机专家开发的程序), 并对您的答案自动进行识别和分析，评估您的作答质量，**

并进行统计排名，对您本次管理能力测评的结果进行评估反馈” (相应展示小 ai 形象, 见网络版附录)。随后, 被试被告知“由于需要等待和评估其他参与者的表现并进行最终排名, 绩效反馈需要约 1 分钟”。最后, 要求被试完成任务类型的操纵检验题目“请问您在多大程度上认为本次您完成的测评题目属于客观任务”, 从 1 = 非常主观至 7 = 非常客观。

在绩效反馈阶段, 参考 Goodman 和 Wood (2004)的做法, 通过为被试解释任务目的, 提供具体的鼓励或改善意见来提升绩效反馈的质量。首先, 要求被试登陆 SalesSmartly (一个专业的企业−人员实时聊天交互网站), 由实验人员扮演的人类管理者或者 AI 为参与者提供几乎实时的绩效反馈。被试通过输入其对应的编号, 即可收到测评的绩效反馈。其次, 在反馈的内容方面, 第一, 实验人员会依据个体作答的详细性、逻辑性以及清晰性, 给予参与者一句测评的总评, 例如: “亲爱的××参与者, 感谢您完成本次管理能力测评的试题。整体上, 您的回答较为模糊(或清晰)”。接着, 向参与者解释测评的目的以及题目考察的具体能力。主观任务组的描述为：“本次测试的第一封公文旨在考察您在建设团队中的矛盾化解能力。而第二封意在测试您应对团队中突发事件的问题解决能力”。客观任务组的描述对应为: “本次测试的第一封公文旨在考察您在原料采购中的运算分析能力。而第二封意在测试您预测销售量的逻辑推理能力”。随后, 依据事先整理好的作答要点对被试作答的两道题目进行评分, 并给予具体的意见。例如: “相较于同期的参与者, 您表现出了较弱的矛盾化解能力【54.15/100】(分析并列举具体的作答缺陷)。但您表现出的突发事件解决能力较好, 得分为【69.75/100】(分析并列举具体的作答优点)”。此外, 为提供明确的负面绩效反馈并更好地对内容进行控制, 所有参与者统一收到: “总的来说, 您在本次测试中的总分为【61.95/100】, 低于 82%的参与者且位于后 18%, 表现有待提升”。为展现人机反馈操纵的差异, 在人类管理者组, 测评专员王亮会感谢被试的参与和配合。在 AI 组, 小 ai 会感谢被试的使用。最后, 为防止产生额外变量, 所有负面绩效反馈均参考相同的格式, 且内容控制在 200 字左右。

阅读反馈后, 被试会填写第二阶段问卷, 为确保被试认真阅读反馈并基于反馈内容填答后续题目, 要求被试回忆并简述收到的反馈内容。接着, 被试填写问卷题目, 包括反馈提供者(1 = 测评专员王亮; 7 = 测评助手小 ai)与反馈内容(1 = 很负面; 5 = 很正面)的操纵检验, 以及对于 AI 的熟悉程度; 绩效改进动机(实验 4 该测量的内部一致性系数为 0.89); 内部归因(内部一致性系数为 0.79)与外部归因(内部一致性系数为 0.83; 测量题项同实验 3)。最后, 与实验 3 类似, 考虑到反馈公平感(内部一致性系数为 0.95)与准确性(内部一致性系数为 0.87)可能作为本研究的替代中介变量, 要求被试分别对上述两个变量进行填答。问卷填答过程中会随机出现两道注意力检测题目(此题请选择“非常不同意”)用于筛选未认真作答的被试。

5.2 结果

5.2.1 操纵检验

首先, 为检验负面绩效反馈的启动效果。要求被试回忆接收的反馈内容: “请问您收到的反馈对您在测评中的表现的评价是 (1 = 很负面, 5 = 很正面)”。结果显示, 个体对反馈内容感知的均值为 2.01(SD = 0.97)。说明实验 4 对负面绩效反馈的操纵是成功的。

其次, 为检验人机提供负面绩效反馈的启动效果。要求被试阅读实验材料后回忆负面绩效反馈的提供者: “请回忆刚才是谁为您提供的绩效反馈(1 = 测评中心负责人王亮, 7 = 测评中心 AI 助手小 ai)”。结果显示, AI 负面绩效反馈组(M = 5.71, SD = 1.72)的评分显著高于人类管理者组(M = 1.60, SD = 1.33), t (158) = 16.92, p < 0.001, Cohen’s d = 0.80。说明实验 4 对人机负面绩效反馈的操纵是成功的。

最后, 为检验任务类型的启动效果, 要求被试评价实验任务的客观性。结果表明, 客观任务组(M = 5.49, SD = 1.23)的任务客观性评分显著高于主观任务组(M = 3.41, SD = 1.51), t (158) = 9.53, p < 0.001, Cohen’s d = 0.60。表明任务类型的操纵成功。

5.2.2 假设检验

为检验研究假设 1。独立样本 t 检验结果表明, 相较于人类管理者(M = 5.76, SD = 0.97), AI 的负面绩效反馈(M = 6.07, SD = 0.78)能够引发更高水平的绩效改进动机, t(158) = 2.24, p = 0.027, Cohen’s d = 0.35。将自我效能感作为协变量控制, 依然发现相较于人类管理者, AI 提供负面绩效反馈会导致更高水平的绩效改进动机, F(1, 157) = 4.98, p = 0.027, η^2 = 0.031。假设 1 得到验证。其次, 为排除 AI 熟悉程度对结果的可能影响, 相关分析发现, AI 组被试对 AI 的熟悉度与其绩效改进动机并不相关(r =

0.058, p = 0.61)。为检验研究假设 2, 双因素方差分析结果发现(如图 3), 人机负面绩效反馈和任务类型对个体的绩效改进动机具有显著的交互作用 F(1, 156) = 44.76, p < 0.001, η^2 = 0.223。且简单效应分析发现, 在客观任务组, AI 负面反馈组的绩效改进动机水平(M = 6.19, SD = 0.72)显著高于人类管理者组(M = 5.09, SD = 0.81), F(1, 156) = 43.66, p < 0.001, η^2 = 0.219。在主观任务组, 人类管理者负面反馈组的绩效改进动机水平(M = 6.43, SD = 0.61)显著高于 AI 组(M = 5.95, SD = 0.82), F(1, 156) = 8.14, p = 0.005, η^2 = 0.05。为检验上述结果的稳健性, 将被试自我效能感作为协变量进行控制, 结果发现, 人机负面绩效反馈与任务类型的交互作用依然显著, F(1, 155) = 44.66, p < 0.001, η^2 = 0.224。因此, 研究的假设 2 再次得到了验证。

为检验内部与外部归因的中介效应, 选择 PROCESS 插件的模型 4, Bootsrap 为 2000。结果表明, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应显著, 且中介效应的指标值为 0.17, 95%的 CI 为[0.017, 0.359], 区间不包含 0。此外, 外部归因在人机负面绩效反馈到绩效改进动机的中介效应指标值为−0.10, 95%的 CI 为[−0.086, 0.011], 区间包含 0, 说明外部归因在人机负面绩效反馈到绩效改进动机的间接效应不显著。研究假设 3 得到了部分验证。

此外, 本研究还检验了内部与外部归因在人机负面绩效反馈与任务类型对绩效改进动机交互效应中的间接效应。选择 PROCESS 插件的模型 8, 设定样本量为 2000。结果表明, 在主观任务中, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应显著, 95%的 CI 为[−0.373, −0.035]; 在客观任务中, 内部归因在人机负面绩效反馈到绩效改进动机的间接效应同样显著, 95%的 CI 为[0.241, 0.669]。且不同任务类型下, 有调节的间接效应的差值显著, 指标值为 0.63, 95%的 CI 为[0.349, 0.989]。说明存在有调节的中介效应; 而外部归因有调节的间接效应指标值为 0.013, 95%的 CI 为[−0.017, 0.110]。区间包含 0, 因此研究假设 4 也得到了部分验证。

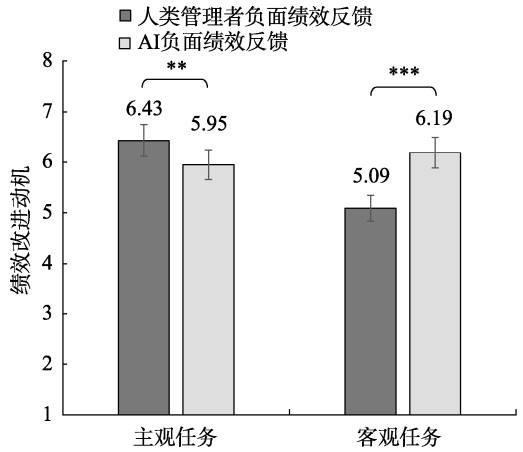

图 3　人机负面绩效反馈与任务类型对绩效改进动机的交互作用图

注：$^{**}p$ < 0.01, $^{***}p$ < 0.001.

为检验反馈准确性和公平感是否能作为替代性解释的机制。首先, 独立样本 t 检验发现, 人类管理者(M = 5.33, SD = 0.91)或 AI 提供负面绩效反馈(M = 5.48, SD = 0.99)在准确性水平上并无差异, t(158) = 0.99, p = 0.32。但是, AI 提供负面绩效反馈(M = 6.14, SD = 1.21)比人类管理者(M = 5.74, SD = 1.28)更公平, t(158) = 2.01, p = 0.046, Cohen's d = 0.33。此外, 人机负面绩效反馈与任务类型对反馈准确性[和公平感]的交互作用均不显著, F(1, 156) = 0.45, p = 0.51[F(1, 156) = 0.20, p = 0.65]。最后, 反馈准确性(指标值为 0.02; 95%的 CI 为[−0.012, 0.105])和公平感(指标值为 0.03; 95%的 CI 为[−0.075, 0.051])在人机负面绩效反馈与绩效改进动机之间的间接效应也均不显著。综上, 同实验 3, 实验 4 也排除了反馈准确性与公平感可能的替代解释机制。

5.3　讨论

实验 4 在实验 3 的基础上, 采用相对真实的反馈策略进一步加强了负面绩效反馈的质量。在为个体提供个性化反馈的基础上, 使得反馈内容更加具体和可信。例如, 相较于实验 3 个体知觉到的反馈准确性(M = 3.79, SD = 1.61), 实验 4 的反馈准确性得到明显的提升(M = 5.41, SD = 0.95)。此外, 实验 4 还进一步验证了先前 3 个实验的结果, 并排除了反馈准确性与公平感可能的替代解释机制。

6　总讨论

基于归因理论, 本研究采用 4 个递进式实验发现了人机负面绩效反馈对绩效改进动机的差异化影响及机制。具体而言, 本研究发现, 相较于人类管理者, 由 AI 提供负面绩效反馈会导致员工更高水平的绩效改进动机。第二, 人机负面绩效反馈和任务类型交互影响个体的绩效改进动机。具体来说, 在主观任务中, 相较于 AI, 个体对人类管理者负面绩效反馈的绩效改进动机更强; 而在客观任务中, 上述结果则发生反转。此外, 本研究基于归因理论,

进一步发现了内部归因在人机负面绩效反馈和任务类型对绩效改进动机的交互作用中起到中介作用。另外, 本研究采用了不同类型的 AI 主体(实验 1 为嵌入式 AI, 实验 2~4 为机器人式 AI)、不同的绩效反馈场景(实验 1~4 分别为电话销售人员、新入职员工培训、员工职业能力竞赛、中层管理者管理能力测评)、差别化的绩效反馈策略(实验 1~3 采用虚假反馈, 实验 4 采用相对真实的反馈), 以及不同的绩效反馈途径(实验 1~2 在线上实验平台展示反馈, 实验 3 通过真实的邮件发送反馈, 实验 4 则通过实时的对话交互网站发送反馈)。总体来看, 本研究 4 个实验的结果保持了较强的一致性和稳健性。

6.1 理论贡献

首先, 本研究拓展了既有负面绩效反馈研究的视角。具体而言, 传统基于人际互动的负面反馈研究大多关注来自人类管理者的反馈(Kitz et al., 2023), 而本研究则发现 AI 替代人类管理者提供负面绩效反馈潜在的积极效应。既有研究从多种角度探索提升负面绩效反馈实施效果的途径, 例如, 反馈特征层面(绩效反馈的频率、即时性或质量等)(Kuvaas et al., 2017; Ni & Zheng, 2024), 员工个体层面(对负面绩效反馈的积极归因, 员工核心自我评价等) (马璐 等, 2021; Xing et al., 2023)。而本研究结合数智化时代背景, 基于人机提供负面绩效反馈的新兴视角, 发现 AI (相较人类管理者)提供的负面绩效反馈提升员工后续的绩效改进动机, 为负面绩效反馈的人机差异化影响效果提供了研究证据。

其次, 本研究丰富了既有人机反馈的研究。当前数智化技术在绩效反馈中的应用引发了一些争论(董毓格 等, 2022)。一方面, 基于算法欣赏的视角, 研究者发现 AI 能够提升绩效反馈的准确性和可靠性, 从而提升员工的绩效水平(Tong et al., 2021), 但另一方面, 也有研究从算法厌恶角度出发, 发现 AI 缺乏真诚性与独特性, 并且会威胁人类的工作机会, 因此当组织披露绩效反馈(尤其是带有鼓励、赞扬性质的正面反馈)来源于 AI 时(Yalcin et al., 2022), 会降低个体的积极表现(Luo et al., 2019; Tong et al., 2021)。本研究聚焦于负面绩效反馈, 并发现 AI (较人类管理者)作为反馈提供者提升个体的绩效改进动机。此外, 既有研究也关注人机反馈产生差异化效果的边界条件, 比如, Tong 等(2021)发现, 对于任期较长的员工而言, 由于他们与组织建立了更强的情感纽带, 对于组织采用 AI 提供绩效反馈的变革也更为支持, 因此员工的任期会缓解 AI 提供绩效反馈的负面效果。此外, Luo 等(2019)发现, 顾客对于 AI 的熟悉程度会降低个体对于 AI 的刻板印象(例如, 缺乏知识和同理心), 从而缓解由 AI 提供反馈造成的产品销量下降。本研究关注员工工作的任务类型这一外部因素, 并发现人机负面绩效反馈与任务类型对员工绩效改进动机的交互作用, 从而拓展了人机反馈边界条件的研究。

此外, 本研究丰富了敏捷型(agile)绩效管理的研究。具体来说, 传统以年、季度为时间单位的绩效管理模式存在周期过长的缺陷, 不利于员工即时获取信息并提升绩效。为此, 有学者提出敏捷型绩效管理的变革趋势(Pulakos et al., 2019; Schleicher et al., 2018), 旨在提升绩效管理的时效性, 并为员工提供准确, 高质量的绩效评估与反馈。数智化是提升敏捷绩效管理最为重要的因素, 表现在 AI 能够不知疲倦地整合并分析数据, 以客观、无偏的方式评估员工的绩效表现, 并提供更加准确的绩效反馈(Qin et al., 2023; Tong et al., 2021)。除人机绩效反馈的研究外, 也有研究关注了 AI 绩效指导(AI coach)。比如, Luo 等(2021)发现 AI 教练相对于人类教练的指导效果在不同的销售人员中呈倒 U 形分布。这是因为绩效排名靠后的销售会面临 AI 反馈信息过载的问题, 而绩效排名靠前的销售对 AI 的厌恶程度较高。本研究与上述文献一致, 均探索了数智化技术对绩效管理中特定环节的影响和机制。

最后, 本研究深化了归因理论在组织场景中的研究。归因理论被广泛应用于解释人际互动中个体如何理解自身或他人行为的原因(Tolli & Schmidt, 2008)。根据经典归因理论的观点(Heider, 1958), 人们通常出于自我防御的目的对不利性结果进行外部归因, 或对有利性结果进行内部归因从而获得自我提升。不过上述结论也受到一些因素的调节, 比如, Xing 等(2023)发现员工核心自我评价水平越高, 越会将负面绩效反馈视作提升与改善绩效的机会, 从而提高内部归因与学习绩效。而本研究深入探究人机反馈的差异化影响, 发现 AI (较人类管理者)提供负面绩效反馈可能会提升个体的内部归因。并结合人机不同的反馈特征(比如, 相较于人类, AI 具备更少的主观或伤害意图) (蒋路远 等, 2022)进行了解释。本研究结果表明, 在削弱负面刺激的消极影响时(如采用 AI 替代人类管理者进行负面绩效反馈), 个体可能会加强内部归因。这为归因理论解释不利性结果中个体的归因倾向或行为表现提供了

新的认识。

6.2 管理启示

本研究也具有一定的管理启示。首先, 传统由人类管理者主导的负面绩效反馈可能破坏领导−下属关系, 引起员工负面情绪并降低绩效水平(Ni & Zheng, 2024)。而本研究的结果表明, AI (较人类管理者)增强了个体的内部归因与绩效改进动机。本研究启示组织可以应用数智化技术赋能绩效管理流程, 发挥 AI 客观、无偏的绩效反馈优势。这一方面能够减轻人类管理者提供负面绩效反馈的压力, 另一方面, 来自 AI 的负面绩效反馈更容易被员工接纳, 从而提升反馈实施的效果。

第二, 尽管数智化技术具有高效、客观、标准化等优势, 但它减少了绩效反馈过程中的人际互动或同理心(董毓格 等, 2022; Yalcin et al., 2022), 因此需要区分人机反馈不同的应用场景。根据本研究的结果, 组织应关注人机负面绩效反馈中的任务特征, 比如, AI 以其客观和无偏的特征为客观任务(比如业绩分析、销量预测等)中的负面绩效反馈提供优势。但相比于人类管理者, 由于 AI 缺乏社会与互动属性, 因而在主观工作任务(比如人际沟通、冲突管理等)中进行负面绩效反馈的效果较差。据此, 组织应事先辨别任务的类型, 从而充分发挥人类管理者与 AI 各自的反馈优势。

第三, 本研究为人机负面绩效反馈后, 组织帮助员工进行积极的心理建设提供了管理启示。由于员工对人机负面绩效反馈的内部归因会影响员工的绩效改进动机, 对负面绩效反馈的内部归因越高, 绩效改进的动机也越高。因此, 组织需要关注人机负面绩效反馈后员工的归因方式, 并加强绩效沟通, 帮助员工及时地发现自身不足或改善绩效反馈流程, 从而提升员工绩效。

6.3 研究局限与展望

本研究也存在一些局限。首先, 未来 AI 可能以人类的外在形象(例如, 虚拟员工)进入工作场所, 与人类员工共事并提供绩效反馈(Yam et al., 2023)。未来研究可以在更为真实的人机反馈场景(例如, 虚拟同事提供反馈), 操控被试的绩效反馈来源感知(来自人类 vs. AI vs.人机混合), 从而加深人机的比较。其次, 以往研究指出, 反馈的特征是影响绩效反馈效果的重要因素。本研究主要关注以绩效表现排名为呈现方式的客观型反馈(objective feedback)。由于人类管理者与 AI 在沟通及情感属性方面的差异, 未来研究可以更多地探索人机在评价型反馈(evaluative feedback; 比如开放性, 质性或针对具体问题的反馈)方面的影响差异(Johnson, 2013)。

另外, 未来研究可以更多关注人机绩效反馈中的文化因素。比如, 在中国传统中庸和谦和文化的影响下, 管理者为避免冲突往往采用“三明治”式的反馈形式, 即在负面的反馈中夹杂鼓励与赞扬性质的正面反馈。由于 AI 通常因缺乏“人情味”而遭到个体的厌恶(Dietvorst et al., 2015; Luo et al., 2019), 未来研究可以探索 AI 采用“三明治”式的绩效反馈策略对员工绩效表现的影响。此外, 相比于西方, 东方社会在和谐文化的影响下, 人际间的沟通方式更为含蓄(耿紫珍 等, 2020), 这可能导致东方社会中的个体更加消极地应对负面绩效反馈, 继而影响人机负面绩效反馈的差异化效果。未来研究可以采用西方样本, 探究文化背景差异下人机负面绩效反馈的效果。

再者, 本研究聚焦于归因理论中的因果控制点视角(内部和外部归因)。事实上, 归因理论的内涵非常丰富。比如, 按照归因的稳定性水平, 个体的归因可被划分为能力归因(ability attribution; 即把事件的结果归因于自身的能力)与努力归因(effort attribution; 即归因于自身的努力或投入程度)。按照归因的可控性, 个体可能会将事件结果归结为可控因素(能力、努力等), 抑或不可控因素(运气、任务难度等) (Russell, 1982; Weiner, 1985)。考虑到 AI 能够基于大数据对个体进行画像性分析, 并深入解析人类在性格、爱好与能力等方面的特征(Fan et al., 2023), 未来研究可以基于更多归因理论的视角, 或进行视角结合(例如, AI 提供的负面绩效反馈能否提升员工对于能力的内部归因, 并影响绩效改进动机), 从而进一步丰富归因理论对人机绩效反馈产生差异化效果的解释。

最后, 本研究关注人机负面绩效反馈影响员工绩效改进动机这一近端的结果。未来研究可以探索人机提供负面绩效反馈对员工实际行为表现(例如绩效水平, 学习行为等)的影响, 从而拓展人机负面绩效反馈的影响后效研究。

参 考 文 献

Audia, P. G., & Locke, E. A. (2003). Benefiting from negative feedback. *Human Resource Management Review, 13*(4), 631−646.

Belschak, F. D., & Den Hartog, D. N. (2009). Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology, 58*(2), 274−303.

Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930−942.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent Algorithm aversion. *Journal of Marketing Research, 56*(5), 809−825.

Chory, R. M., & Westerman, C. Y. (2009). Feedback and fairness: The relationship between negative performance feedback and organizational justice. *Western Journal of Communication, 73*(2), 157−181.

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618−630.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114−126.

Dimotakis, N., Mitchell, D., & Maurer, T. (2017). Positive and negative assessment center feedback in relation to development self-efficacy, feedback seeking, and promotion. *Journal of Applied Psychology, 102*(11), 1514−1527

Dong, Y., Long, L., & Cheng, Z. (2022). Performance management in the era of digital intelligence: Present and future. *Tsinghua Management Review, 5*, 93−100.

[董毓格, 龙立荣, 程芷汀. (2022). 数智时代的绩效管理：现实和未来. *清华管理评论, 5*, 93−100.]

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology, 108*(8), 1277−1299.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175−191.

Fields, B., Carbery, M., Schulz, R., Rodakowski, J., Terhorst, L., & Still, C. (2023). Evaluation of Face Validity and Acceptability of the Care Partner Hospital Assessment Tool. *Innovation in Aging, 7*(2), 557−568.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

Geng, Z. Z., Zhao, J. J., & Ding. L. (2020). The “wisdom” of golden mean thinking: Research on the mechanism between supervisor developmental feedback and employee creativity. *Nankai Management Review, 23*(1), 75−86.

[耿紫珍, 赵佳佳, 丁琳. (2020). 中庸的智慧：上级发展性反馈影响员工创造力的机理研究. *南开管理评论, 23*(1), 75−86.]

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals, 14*(2), 627−660.

Goodman, J. S., & Wood, R. E. (2004). Feedback specificity, learning opportunities, and learning. *Journal of Applied Psychology, 89*(5), 809−821.

Hareli, S., & Hess, U. (2008). The role of causal attribution in hurt feelings and related social emotions elicited in reaction to others' feedback about failure. *Cognition and Emotion, 22*(5), 862−880.

Harvey, P., Madison, K., Martinko, M., Crook, T. R., & Crook, T. A. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives, 28*(2), 128−146.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons Publishing House.

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review, 39*, 105456.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349−374.

Jiang, L., Cao, L., Qin, X., Tan, L., Chen, C., & Peng, X. (2022). Fairness perceptions of artificial intelligence decision-making. *Advances in Psychological Science, 30*(5), 1078−1092.

[蒋路远, 曹李梅, 秦昕, 谭玲, 陈晨, 彭小斐. (2022). 人工智能决策的公平感知. *心理科学进展, 30*(5), 1078−1092.]

Johnson, D. A. (2013). A component analysis of the impact of evaluative and objective feedback on performance. *Journal of Organizational Behavior Management, 33*(2), 89−103.

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal, 63*(2), 584−612.

Kitz, C. C., Barclay, L. J., & Breitsohl, H. (2023). The delivery of bad news: An integrative review and path forward. *Human Management Review, 33*(3), 1−23.

Kluger, A. N., & Denisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254−284.

Kuvaas, B., Buch, R., & Dysvik, A. (2017). Constructive supervisor feedback is not sufficient: Immediacy and frequency is essential. *Human Resource Management, 56*(3), 519−531.

Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the "More is better" assumption. O*rganizational Behavior and Human Decision Processes, 116*(2), 217−228.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society, 5*(1), 1−16.

Leo, X., & Huh, Y. E. (2020). Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Computers in Human Behavior, 113*, 106520.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629−650.

Luo, X., Qin, M. S., Fang, Z., & Qu, Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing, 85*(2), 14−32.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science, 38*(6), 937−947.

Lyytinen, K., Nickerson, J. V., & King, J. L. (2021). Metahuman systems = humans + machines that learn. *Journal of Information Technology*, *36*(4), 427−445.

Ma, L., Xie, P., Wei, Y., & Qiao, X. (2021). Is negative feedback from leaders really harm for employee innovative behavior? The role of positive attribution and job crafting. S*cience and Technology Progress and Policy, 38*(12), 144−150.

[马璐, 谢鹏, 韦依依, 乔小涛. (2021). 领导者负面反馈真的不利于员工创新吗——积极归因与工作重塑的作用.

科技进步与对策, *38*(12), 144−150.]

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149−167.

Ni, D., & Zheng, X. M. (2024). Does negative performance feedback always lead to negative responses? The role of trust in the leader. *Journal of Occupational and Organizational Psychology*, *97*(2), *623−646*.

Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*(1), 45−67.

Pulakos, E. D., Mueller-Hanson, R. A., & Arad, S. (2019). The evolution of performance management: Searching for value. *Annual Review of Organizational Psychology and Organizational Behavior, 6*, 249−271.

Qin, S., Jia, N., Luo, X., Liao, C., & Huang, Z. (2023). Perceived fairness of human managers compared with artificial intelligence in employee performance evaluation. *Journal of Management Information Systems, 40*(4), 1039−1070.

Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes, 164*, 11−26.

Russell, D. (1982). The causal dimension scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology, 42*(6), 1137−1145.

Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management, 44*(6), 2209−2245.

Scholz, U., Gutiérrez-Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242−251.

Song, X., & He, X. (2020). The effect of artificial intelligence pricing on consumers' perceived price fairness. *Journal of Management Science, 33*(5), 3−16.

[宋晓兵, 何夏楠. (2020). 人工智能定价对消费者价格公平感知的影响. *管理科学*, *33*(5), 3−16.]

Tolli, A. P., & Schmidt, A. M. (2008). The role of feedback, casual attributions, and self-efficacy in goal revision. *Journal of Applied Psychology, 93*(3), 692−701.

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal, 42*(9), 1600−1631.

Van Dijk, D., & Kluger, A. N. (2011). Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective. *Journal of Organizational Behavior, 32*(8), 1084−1105.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548−573.

Wexley, K. N., Singh, U. A., & Yukl, G. A. (1973). Subordinate personality as a moderator of the effects of participation in 3 types of appraisal interviews. *Journal of Applied Psychology, 58*(1), 54−59.

Xing, L., Sun, J. M., Jepsen, D., & Zhang, Y. J. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 76*(2), 1−31.

Xu, L., Yu, F., & Peng, K. (2022). Algorithmic discrimination causes less desire for moral punishment than human discrimination. *Acta Psychologica Sinica, 54*(9), 1076−1092.

[许丽颖, 喻丰, 彭凯平. (2022). 算法歧视比人类歧视引起更少道德惩罚欲. *心理学报, 54*(9), 1076−1092.]

Yalcin, G., Lim, S., Puntoni, S., & van Osselaer, S. M. J. (2022). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696−717.

Yam, K., Tang, P., Jackson, J., Su, R., & Gray, K. (2023). The rise of robots increases job insecurity and maladaptive workplace behaviors: Multimethod evidence. *Journal of Appied Psychology, 108*(5), 850−870.

A comparative study on human or AI delivering negative performance feedback influencing employees' motivation to improve performance

WANG Guoxuan[1], LONG Lirong[1], LI Shaolong[2], SUN Fang[3], WANG Jiaqing[1], HUANG Shiyingzi[1]

([1] *School of Management, Huazhong University of Science and Technology, Wuhan* 430074, *China*)

([2] *Economics and Management School, Wuhan University, Wuhan* 430072, *China*)

([3] *Business School, Hubei University of Economics, Wuhan* 430205, *China*)

Abstract

Given that negative performance feedback can evoke negative reactions from employees, delivering such feedback effectively has become a challenge for organizations. Driven by rapid innovations in science and technology, artificial intelligence (AI) is gradually being applied in organizational management. For instance, AI can monitor employees' work behaviors in real time, diagnose and analyze the reasons for their poor performance, and provide them with suggestions for performance improvement. Based on attribution theory, this study examined the benefits that employees may gain when receiving negative performance feedback from AI

than from human managers. This study also explored the moderating effect of task type (subjective vs. objective) as a boundary factor for the influence of feedback source on employees' motivation to improve their performance. Previous research has shown that individuals often experience internal or external attribution after receiving negative performance feedback. Therefore, this study also proposed internal and external attributions as underlying mediating mechanisms.

To test the hypotheses, four experiments were conducted (N = 598) involving various kinds of performance feedback contexts in the workplace, including performance feedback for employees in call centers, training of new employees, employees in vocational ability competitions, and middle managers' capacity of management. Two strategies were adopted to provide performance feedback for the participants. Specifically, experiments 1–3 used fake feedback to control the feedback content among the participants, while experiment 4 delivered relatively real negative feedback based on the participants' actual performance to further test the results of experiments 1–3.

Experiment 1 involved 128 full-time employees and used a single-factor, two-level between-subjects design, and results showed that compared with those coming from human managers, negative performance feedback coming from AI led to a higher employee motivation to improve performance. Experiment 2 involved 160 employees and used a two-factor between-subjects design, and results highlighted the interactive effect of negative feedback source (human manager vs. AI) and task types on employees' motivation to improve their performance. Specifically, in subjective tasks, negative performance feedback from human managers (relative to AI) resulted in a higher motivation to improve performance. However, the opposite case was observed in objective tasks. Experiment 3 involved 150 employees who received negative performance feedback through email, and results highlighted the mediating role of internal attributions in the relationship between negative feedback source and motivation to improve performance. Experiment 4 involved 160 employees and utilized relatively real negative performance feedback, and results were the same as those obtained in experiment 3.

This study offers four theoretical contributions. First, with the emergence of AI as a feedback source for organizations, results show that compared with feedback from human managers, negative performance feedback from AI led to a higher motivation among employees to improve their performance, thereby enriching traditional research on negative performance feedback. Second, task type can moderate the relationship between negative performance feedback source and employees' motivation to improve performance, and this finding contributes to the expansion of the boundary effects of feedback source (AI or human managers). Third, this study generates insights into agile performance management in the digital intelligence era by demonstrating the advantages of AI in replacing human managers in delivering negative performance feedback. Fourth, this study underscores internal attribution as a potential mechanism, thus expanding the application of attribution theory in explaining individual motivation and behavior within the context of AI versus human managers in delivering negative performance feedback.

Keywords negative performance feedback, artificial intelligence, motivation to improve performance, internal and external attribution, task type

附录: 主要实验材料

1. 附录 1 (自编主观和客观任务)

注：实验 2 采用主观任务 1 与客观任务 3; 实验 3 采用以下所有任务; 实验 4 采用主观任务 1、2, 以及客观任务 2、3。

[主观任务]

假设今天是××年×月×日, 您是×公司的××。现在您收到了三封公文邮件需要您处理。

[主观任务 1]

您收到了××部门同事××给您发送的邮件公文, 公文中提到了××在带领团队完成项目过程中遇到的困难, 请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于项目团队建设问题

××：

最近, 公司派我带领团队负责一个新项目。我们项目团队结合以往经验顺利完成了前期工作, 也得到了甲方的确认和认可。但由于时间紧任务重, 又申请从公司调来了 2 名新成员进入项目团队。项目开始实施后, 项目团队原成员和新加入的成员之间经常发生争执, 对出现的错误相互推诿。项目团队原成员认为新加入的成员效率低下, 延误了项目进度; 新加入成员则认为项目团队原成员不好相处, 没法有效沟通。我开始认为这是正常的团队磨合过程, 没有过多的干预。但是项目实施 2 个月之后, 我发现这样下去肯定会出问题。对此, 您有何建议？请指示。

××部门××

××年×月×日

问题：请您结合自己的工作经验，给××部门同事××提供有效的建议(建议要具体可行，逻辑清晰，不少于 100 字)。

[主观任务 2]

您收到了××部同事××给您发送的邮件公文, 公文中提到了××在带领团队完成项目过程中遇到的困难, 请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于突发情况的处理

××：

目前, 我主要负责的一个项目正进行到后期阶段, 但最近因为疫情的突然反复, 原本负责一项线下工作的项目同事小李被隔离在外。由于成本的控制, 我们项目团队今年已不能再招收新人, 而该线下工作之前一直只有小李主要负责, 其他人并不了解且自身工作相对饱和。受条件限制, 该工作也不能线上进行。对此, 您有何建议？请指示。

××部门××

××年×月×日

问题：请您结合自己的工作经验，给××部门同事××提供有效的建议(建议要具体可行，逻辑清晰，不少于 100 字)。

[主观任务 3]

您是××的职场导师，他给您发送了邮件公文，公文中提到了其在近日工作中遇到的困难，请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于工作沟通的求助

师傅：

今天，总经理安排我和××一起完成一项重要的工作任务。当我询问领导何时需要交付工作时，领导回答尽快完成。原本，我打算今天加班完成该工作任务，但××认为领导并没有说何时交付工作，可以下周再慢慢完成。我劝说了很久，他还是觉得没必要加班。请问我应该如何劝说他和我一起加班完成这项工作任务？您能否给我一些建议？请指示。

徒弟××

××年×月×日

问题：请您结合自己的工作经验，给您的职场徒弟××提供有效的建议。(要具体可行，逻辑清晰，不少于 100 字)

[客观任务]

假设今天是××年×月×日，您是××公司××部门××。现在您收到了三封公文邮件需要您处理。

[客观任务 1]

您收到了××部门同事××给您发送的邮件公文，公文中提到了关于工作安排中的问题，请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于工作安排问题

××：

明天，我们邀请了甲、乙、丙、丁 4 名应聘者同时来公司项目部参加三个阶段的面试，公司要求每名应聘者都必须先由招聘专员进行初试，然后到项目主管处进行复试，最后到项目部经理处参加面试，并且不允许插队(即：在任何一个阶段 4 名应聘者的顺序是一样的)，由于 4 名应聘者的专业背景和工作经历不同，所以每人在三个阶段的面试时间也不同，如下表所示：(单位：分钟)

	招聘专员初试	项目主管复试	项目部经理面试
甲	13	15	20
乙	10	20	18
丙	20	16	10
丁	8	10	15

现在，由于我们公司距离应聘者下榻的酒店较远，因此要安排专车送 4 名应聘者一起同时离开公司。明天早晨 8: 00 开始面试，请问您怎样安排他们的面试顺序，可以让他们一起最早离开公司？最早时间是几点？请您具体安排。

××部门××

××年×月×日

问题：请您运用逻辑推理和运算能力，回复××关于面试顺序安排的问题，对甲、乙、丙、丁的面试顺序进行排序，并推算出他们能够离开公司的最早时间。(请写出具体的逻辑分析和推理，不少于 100 字)

[客观任务 2]

您收到了××部门同事××给您发送的邮件公文, 公文中提到了项目生产设备的投资购买问题, 请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于生产设备的投资购买

××：

我们项目团队最近预计投资购买一个项目生产设备, 并按 6 个决策指标对不同类型的设备进行综合评价。这 6 个指标是, 最大运行速度(C1)、最大产量(C2)、最大负荷(C3)、费用(C4)、可靠性(C5)、灵敏性(C6)。现有 4 种型号的设备可供选择, 具体指标值如表。各属性所占比重分别为 20%、10%、10%、10%、20%、30%, 请您指示我们应该如何进行购买？请您就购买设备问题进行决策, 选择最优的购买方案。

指标设备类型	最大运行速度(马赫)	最大产量(件/天)	最大负荷(千克)	费用(万元)	可靠性程度(评估分值)	灵敏性程度(评估分值)
A	2.0	1500	20000	5.5	5	9
B	2.5	2700	18000	6.5	3	5
C	1.8	2000	21000	4.5	7	7
D	2.2	1800	20000	5.0	5	5

××部门××

××年×月×日

问题：请您运用逻辑推理和运算能力，回复××关于设备投资购买的问题，将 A, B, C, D 四种设备按照优劣进行排序。(请写出具体的逻辑分析和推理，不少于 100 字)

[客观任务 3]

您是××的职场导师, 他给您发送了邮件公文, 公文中提到了其最近在项目前期调研中遇到的困难, 请您处理以下公文。

[公文]

公文类型：邮件

发至：××

发自：××

日期：××年×月×日

主题：关于项目调研的求助

师傅：

最近, 我正在为我负责的项目做前期调研, 需要了解该项目涉及到的一商品销售情况以决定是否将该商品纳入我们的项目方案。下表是该商品最近 10 个月的商品销售量统计记录。按公司销量为标准进行划分：(1)销售量<60 千件：属于滞销; (2)60 千件≤销售量≤100 千件：属于一般; (3)销售量 > 100 千件：属于畅销。根据公司要求, 如果该商品下个月(第 11 个月)畅销则可以纳入我们的项目方案。请您指示是否要将该商品纳入我们的项目方案。

商品销量统计表				单位：千件						
时间	1	2	3	4	5	6	7	8	9	10
销售量	40	45	80	120	110	38	40	50	62	90

徒弟-××

××年×月×日

问题：请您运用逻辑推理和运算能力，回复××关于项目调研的求助问题，分析第 11 个月商品销量状态(滞销、一般、畅销)，并预测第 11 个月的商品销售量。(请写出具体的逻辑分析和推理，不少于 100 字)_____

2. 附录 2(展示人类管理者或 AI 形象的图片)

实验 1 销售经理小艾与算法系统 ai

实验 2~4 测评专员王亮与人工智能测评助手小 ai